文献与数据学报
第 6 卷　第 4 期　　2024 年 12 月

# 守护文化根脉：新时代语言资源建设的思考

张永伟
（中国社会科学院语言学重点实验室，北京　102488）

DOI：10.31193/SSAP.J.ISSN.2096-6695.2024.04.03

中华优秀传统文化是中华民族的文化根脉。中华民族拥有悠久而辉煌的历史，孕育出了源远流长、博大精深的中华优秀传统文化。文化建设是统筹推进“五位一体”总体布局、协调推进“四个全面”战略布局的重要内容，是推动高质量发展的重要支点，是满足人民日益增长的美好生活需要的重要因素，也是战胜各种风险挑战的重要力量源泉。党的十九届五中全会提出，到 2035 年“建成文化强国”和“国家文化软实力显著增强”的远景目标。文化建设一直备受关注，党的二十届三中全会通过的《中共中央关于进一步全面深化改革、推进中国式现代化的决定》进一步强调，聚焦建设社会主义文化强国，坚持马克思主义在意识形态领域指导地位的根本制度，健全文化事业、文化产业发展体制机制，推动文化繁荣，丰富人民精神文化生活，提升国家文化软实力和中华文化影响力。

作为人类最重要的交际工具，语言不仅承载着民族的历史记忆，也蕴含着特定的文化价值。在文化传承与发展的诸多要素中，语言资源是基石。

## 1　语言资源的基本内涵

语言资源概念的提出已有四十余年，并受到国内学界的长期关注。例如，邱质朴教授早在 1981 年就发表了《试论语言资源的开发——兼论汉语面向世界问题》一文，探讨了语言资源的概念、开发技术、开发与推广以及在汉语推广中的应用等议题。尽管学界长期关注语言资源，但其内涵丰富且不断发展，尚未达成共识。一般而言，狭义的语言资源指在历史演进过程中形成的语言文字总和，包括标准语、方言和少数民族语言在内的书面和口头语言形式。广义的语言资源还涵盖了语言相关的知识、技术、工具、人才等衍生资源。本文讨论聚焦于狭义的语言资源。在全球化和信息化的推动下，语言资源呈现出数字化、集成化和智能化的发展趋势。

## 2　语言资源的文化价值

语言资源作为文化传承的重要载体，在中华优秀传统文化的保存与传播中发挥着不可替代的作用。

［作者简介］张永伟（ORCID：0000-0003-4129-7640），男，研究员，研究方向为语料库语言学、计算语言学、计算词典学，Email：zhangyw@cass.org.cn。

首先，语言资源是文化传承与保护的基础性要素。语言承载着民族的历史记忆和文化基因，蕴含着丰富的民族智慧。通过系统性收集整理和保护语言资源，特别是运用数字化技术构建语言资源数据库，不仅能有效传承民族文化精髓，保存文化遗产，还能为文化研究提供实证素材，有助于揭示文化发展轨迹，深化对民族文化的系统性认知。

其次，语言资源在增强民族凝聚力与构建国家认同方面发挥着基础性作用。共同的语言系统构成了民族群体思维与交际的基础平台，有助于形成文化共识与价值认同。通过语言资源的系统性整理与深入研究，可强化民族文化特质，增强文化自信，培育归属感与认同感。同时，语言资源的标准化建设与规范应用有助于巩固民族凝聚力、维护国家文化主权。

再次，语言资源的有效开发利用是提升国家文化软实力的重要途径。语言资源的创新性开发能推动文化创意产业发展，丰富文化产品供给。数字化和智能化技术的应用能够提升文化传播效能，以适应现代传播需求。此外，语言资源研究成果可为国家文化政策制定提供科学依据，助力国家文化战略的实施。

最后，语言资源在促进跨文化交流中发挥着桥梁作用。通过语言翻译和语言资源共享，可以增进不同文化之间的相互理解，促进文化传播，减少文化交流中的误解和偏见。跨语言的文化资源整合和合作研究有助于推动文明对话，构建多元文化共生格局，丰富世界文明的多样性。

## 3　新时代语言资源建设面临的挑战

我国语言资源建设虽取得显著成就，但在新时代背景下仍面临诸多挑战，以下举例说明。

首先，古代语言资源的系统化建设亟待加强。与现代语言资源的多样化、高度数字化和处理便捷相比，古代语言资源建设存在多重制约因素，如原始资料珍贵稀缺、研究经费不足、学术门槛较高以及专业人才匮乏等。尽管国家已出台支持甲骨学、简牍学、敦煌学、古文字学等“绝学”和冷门学科发展的政策措施，但在资源收集整理的广度和深度方面仍有欠缺。具体表现为古代语言资源数字化程度低，缺乏系统完备的检索数据库，甚至连快速并准确追溯某个字词的出处及用法演变这样的基本要求都无法满足，导致一些基础性的学术考证工作难以开展。

其次，方言及少数民族语言资源建设亟须完善。与体系较为完善的汉语普通话资源相比，方言及少数民族语言资源建设面临更大挑战。由于方言种类繁多、地域分布广泛且部分方言使用人群萎缩，加之少数民族语言中存在大量无文字语言，导致他们的资源收集整理以及数字化难度较大。虽然“中国语言资源保护工程”等重大项目在方言和少数民族语言的保护方面取得一定成效，但要实现全面覆盖、深度记录和高度数字化仍需持续努力。例如，许多方言和少数民族语言缺乏完整的音系描写和系统的语料库，难以支撑深入的语言研究和文化传承。

最后，多媒体多模态语言资源建设亟待推进。目前，基于拼写文本的语言资源已积累海量数据，并形成了较为成熟的处理技术和应用平台。然而，包含语音、图像、视频等模态，甚至涵盖人类表情、姿态、生理信号等信息的多媒体多模态语言资源尚处于起步阶段。其数据采集和标注成本高昂、技术难度大，且缺乏统一的标准规范，导致数据共享和利用效率低下。此外，多媒体多模态数据的集成和融合也存在技术瓶颈，影响了人工智能在情感识别、人机交互、虚拟现实等领域的应用发展。

## 4 语言资源建设的建议

在新时代，语言资源建设要有新起点、新要求、新气象和新作为。鉴于当前语言资源建设面临的挑战，本文提出以下建议。

首先，整合资源，融合发展。在借鉴现代汉语及拼写文本语言资源建设经验的基础上，系统补充其他语言资源建设的不足。同时，全面推进古代语言资源、方言及少数民族语言资源的数字化建设和整合集成，构建统一的应用平台，以促进不同语言资源的协同研究，发挥资源整合的叠加效应。

其次，协同创新，技术赋能。积极推动大数据、人工智能等新兴技术在语言资源建设中的应用。通过技术手段深度挖掘语言资源内蕴的语言规律、文化内涵和社会变迁，并开展多维度的切分标注和智能分析，提升语言资源的整体水平和实用价值。此外，建议在创新性强的跨学科项目申报环节适度降低前期成果要求，鼓励不同领域研究者勇于参与，营造良好的跨学科合作氛围，推动语言资源建设的创新发展。

再次，健全机制，持续投入。为确保语言资源建设的有效性和可持续性，采取差异化投入策略。对于具有基础性地位或文化价值显著但商业价值有限的语言资源，建议将其纳入国家重大科技基础设施建设规划，由国家权威机构牵头，保障其建设质量和长期稳定发展。其他类型的语言资源则应探索多元化、可持续的投入机制，鼓励社会资本参与，不断丰富语言资源的多样性和规模。

最后，优化评价，人才强基。一方面，加强复合型专业人才培养，注重理论与实践的有机结合，推动高校、科研机构、文博机构及企业等多方合作培养专业人才。另一方面，构建科学合理的人才评价机制，破除“五唯”（唯论文、唯帽子、唯职称、唯学历、唯奖项）的评价模式，建立更多元、更全面的评价标准，兼顾个人学术成果与团队协作中的贡献，以激发人才队伍的创新活力。

语言资源建设是国家文化发展战略的重要组成部分。在新时代背景下，应当充分认识语言资源的文化价值，统筹推进语言资源建设，坚持保护传承与开发利用并重的发展理念，不断提升语言资源服务国家文化发展战略的能力，为中华优秀传统文化传承发展和文化强国建设提供坚实支撑。

# 数字技术助力考古学研究

刘建国
（中国社会科学院考古研究所，北京 100101）

**DOI：**10.31193/SSAP.J.ISSN.2096-6695.2024.04.04

数字技术大潮迅猛发展，早已渗透到社会生活的各个方面。考古工作中，数字技术贯穿空间信息的获取、存档、分析、研究、展示、传承等全过程，不断拓展和完善现代考古学研究的领域。

［作者简介］刘建国，研究员，研究方向为数字考古、史前治水文明，Email：jgliu@cass.org.cn。